CLAIMS

1. A method for determining the quality of a result of a clustering data processing operation, the result comprising a set of clusters, a cluster having a set of buckets for each variable, the method comprising the steps of:

a) determining a foreground frequency of a bucket within a first cluster;

b) determining a background frequency of the bucket with respect to all of the clusters;

c) comparing the foreground and background frequencies; and

d) determining a quality index based on the comparison.

2. The method of Claim 1, wherein said comparing step further comprises subtracting the relative foreground and background frequencies.

3. The method of Claim 2, wherein said comprising step further comprises squaring the result of the comparison.

4. The method of Claim 1, further comprising the steps of:

e) determining an optimal number of clusters; and

f) comparing the optimal number of clusters to the actual number of clusters resulting from the clustering date processing operation

5. The method of Claim 4, wherein the optimal number of clusters is determined by a maximum number of buckets for a variable.

6. The method of Claim 5, wherein the optimal number of clusters is set to a threshold value in case the maximum number of buckets is greater than the threshold value.

7. The method of Claim 4, further comprising the steps of:

g) determining a factor based on the optimal number of clusters and the actual number of clusters; and

h) multiplying the result of the comparison of the relative foreground and background frequencies with the factor.

8. The method of Claim 7, further comprising the steps of:

i) determining a normalizing value being independent of any correlations between fields of the data on which the data processing operation is applied; and

j) normalizing the result of the comparison of the foreground and background frequencies by means of the normalizing value.

9. The method of Claim 8, wherein said step of determining the normalizing value further comprises:

i) comparing the background frequencies of the buckets with an imaginary cluster having a foreground frequency of the bucket equal to one;

ii) comparing the background frequencies of the buckets with an imaginary cluster having a foreground frequency of the bucket equal to zero; and

iii) summing the results of the corresponding comparison values.

10. A method for data clustering, said method comprising the steps of:

a) performing a number of data clustering operations;

b) determining a quality index for each result of the data clustering operations; and

c) selecting the result with the highest quality index as an end result of the data clustering.

11. A method for data clustering, said method comprising the steps of:

a) selecting an initial set of clusters;

b) determining a quality index for the clusters; and

c) performing a number of iterations to improve the quality index.

12. The method of Claim 11, further comprising the steps of:

d) moving at least one record of at least one of the clusters to another cluster;

e) determining the quality index for the modified clusters; and

f) using the modified clusters as a new initial set of clusters in case the quality index improved.

13. A computer program product stored on a computer usable medium for determining the quality of a result of a clustering data processing operation, the result comprising a set of clusters, a cluster having a set of buckets for each variable, the method comprising the said program product comprising:

determining first subprocesses for a foreground frequency of a bucket within a first cluster;

determining second subprocesses for a background frequency of the bucket with respect to all of the clusters;

comparing third subprocesses the foreground and background frequencies; and

determining fourth subprocesses a quality index based on the comparison.